DOCUMENT RESUME

ED 483 029 TM 035 225

AUTHOR	Baker, Eva L.
TITLE	From Usable to Useful Assessment Knowledge: A Design Problem. CSE Report.
INSTITUTION	National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.; California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	Institute of Education Sciences (ED), Washington, DC.
REPORT NO	CSE-R-612
PUB DATE	2003-10-00
NOTE	19p.; Keynote presentation at the Annual International Congress for School Effectiveness and Improvement (Sydney, Australia, 2003).
CONTRACT	R305B960002
AVAILABLE FROM	For full text: http://www.cresst.org/reports/ R612.pdf.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Design; *Evaluation; *Evaluation Research; Program Design

ABSTRACT

Policymakers are fond of saying that we have enough research knowledge; however, one problem is that nobody is applying the knowledge we already have. This paper offers a discussion of the kinds of knowledge needed to improve students' learning, what to do about this in the assessment and testing arena, and where assessment falls short. Types of knowledge are discussed, including research knowledge, which must be both usable and useful. Distinctions are made between usable and useful knowledge, and a case is made for how we might ultimately design our systems and our own actions to help us act with greater intelligence. (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

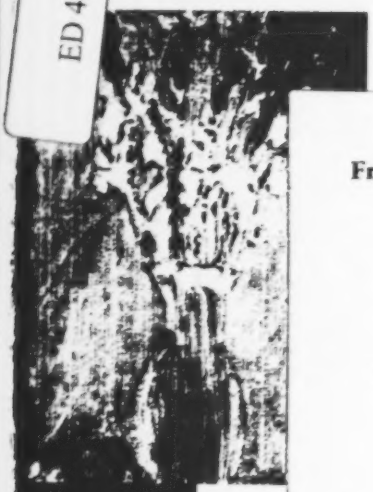
CRESST

National Center for Research on Evaluation, Standards, and Student Testing

1m

ED 483 029

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.



From Usable to Useful Assessment Knowledge: A Design Problem

CSE Report 612

Eva L. Baker

National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
University of California, Los Angeles

TM035225

UCLA Center for the Study of Evaluation

In Collaboration With:

University of Colorado at Boulder • Stanford University • The RAND Corporation
University of Southern California • Educational Testing Service
University of Pittsburgh • University of Cambridge

BEST COPY AVAILABLE

From Usable to Useful Assessment Knowledge: A Design Problem

CSE Report 612

Eva L. Baker

National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
University of California, Los Angeles

October 2003

Center for the Study of Evaluation
National Center for Research on Evaluation, Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Project 2.1 Cognitively Based Models for Assessment and Instructional Design. Strand 1: Cognitively Based Models and Assessment Design
Project Director: Eva L. Baker, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles

Copyright © 2003 The Regents of the University of California

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

FROM USABLE TO USEFUL ASSESSMENT KNOWLEDGE: A DESIGN PROBLEM*

Eva L. Baker

National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
University of California, Los Angeles

Policymakers are fond of saying that we have enough research knowledge. The problem is that nobody is applying the knowledge we already have. In this paper I plan to discuss the kinds of knowledge we need to improve students' learning and, in particular, what to do about it in the assessment and testing arena. I will start with a consideration of knowledge that is supposed to be used, and where assessment falls short. Then I will provide solutions, drawn from my own research. There are types of knowledge that are characterized as basic or fundamental, that may or may not be applied in the foreseeable future. I am interested in education, where application of new knowledge is vitally needed and can't be put off for the next—or the next—generation of 10-year-olds. If research knowledge is to be applied in a reasonable period of time, say one quarter of a lifetime, it should possess certain properties. The knowledge must be both *usable* and *useful*. I'll spend a little time on these distinctions and start with the negative. My intention is to make a case for how we might ultimately design our systems and our own actions to help us act with greater intelligence.

Each of us knows what *un*usable knowledge is. It includes, but is not limited to, names of colleagues unexpectedly encountered at conferences; comments that, while accurate, are far too impolite to utter; random information we know but can't remember why; and fractions of ideas that we can't seem to mold into a functional whole. The most frustrating variant of unusable knowledge is the one in which we know part or all that should be done, having goals valuable to attain, but remain nonetheless unable to make significant progress applying the forms of knowledge that we have at our disposal. This type of unusable knowledge seems to pertain to many aspects of educational improvement.

---

*Keynote presentation at the 2003 International Congress for School Effectiveness and Improvement, Sydney, Australia.

"Usable knowledge" was a term used by Lindblom and Cohen (1979) in a book that focused on the relationship of the social sciences and problem solving in the real world. If you see the actual book, you will notice that in the title, the words "usable knowledge" are not capitalized, either a sign of the authors' sense of its rarity or a tribute to e. e. cummings. Remember, Lindblom and Cohen were specifically considering research and evaluation knowledge, and they found it wanting. Rarely could they find instances that were usable for the specific social problems at hand. Sound familiar?

So what kind of knowledge is valuable for our situations? We will exclude obviously wrong information. I think it may be helpful to make the distinction between *usable* and *useful* knowledge. For knowledge *to be usable*, it needs to be understood and then translated into practical terms so that it has the potential to be applied to a particular situation or problem. *To be useful*, knowledge permits us to act by changing the problem into a solvable form, leading to the development of a solution or greater insight—the difference is between potency and action, a revered St. Thomas Aquinas formulation.

For a somewhat more recent reference, Carol Weiss (1977) discussed how research and evaluation findings can be usable, even though they may not directly lead to a decision about a given situation, and can illuminate the problem so that new approaches are possible. Illumination literally means that we see the problem in a new light and are enabled to reconceive it. For a homely example, we know that it takes many measures and complex equations in order to forecast the weather. If they were not transformed, most of us would have no idea what they meant. Yet, this complex knowledge can be translated into an understandable, usable form, such as "Get ready for rain." If there were no rain, the information would have been usable, but not useful. So the best kind of knowledge would be first usable (able to be translated into relevant terms) and then useful, to help us to act, to demonstrably improve our situation.

Usable and Useful Knowledge in School Reform

Why are some schools or institutions successful in finding knowledge and making it both usable and useful? I believe, besides luck, that the people in these organizations exhibit certain recurring predispositions that add up to efficacy. First, they focus on their primary business: In schools, that is learning of all sorts, by students and educators. Second, they embrace information of both formal and

informal types, finding ways to integrate sources of information. Third, their staffs make information public and exchangeable. And last, they take pride in the outcomes they achieve.

The remainder of my brief remarks will link the concepts of usable and useful knowledge to a specific feature of school reform: assessment—that is, the testing of children or other students, for the purposes of certification, instructional improvement, system monitoring, the evaluation of educational services, and in some cases, the accountability consequences that follow. How does and how could assessment knowledge play out in a school or training environment?

Assessment, the Ever-New Solution

Without question, policymakers at all levels, worldwide, share the belief that assessment knowledge will help them and people in the educational systems to solve problems of teaching, learning, and management. The interest in assessment preceded and extends far beyond its use in education. Assessment suffuses the fields of health policy, transportation, criminal justice, and social services. Because assessment results (how well individuals score on tests) are often decisive in determining access to an improved economic status, it is easy to see why regions, states, and countries have ascribed the same power to test scores—a way to project winners and losers. It is not lost on us that both the private and government sectors have picked up the assessment talisman, for situations ranging from matters of convenience to those of survival. For instance, we rate restaurant cleanliness, nursing homes, quality of intensive care in major hospitals, and environmental factors such as sunshine and amount of pollution. No one in Los Angeles would dream of going to a restaurant with less than an "A" rating, even if the air, inside or out, were unbreathable. We pay attention to one rating because it seems to be related to actions people take. It is useful. In the case of smog, we have a greater challenge. We have usable (the pollution rating) but not really useful knowledge.

For those directly interested in education and training, among policymakers and planners, the level of enthusiasm for assessment is palpable. Their interests include precollegiate schools, early education, university and college education, formal preparation for the workplace, training and development in business, the military, and other technical sectors. Why? Managers have well learned a number of lessons. Assessment information provides a quantitative measure that allows them to make distinctions among people and organizations. The differences can be easily

summarized. They can, they believe, distinguish between the better and the best. They use the classifications that they create (novice, proficient, expert) as a convenient way to allocate services. The assessment procedures they have adopted are but an inexpensive fraction of the cost of total services, and usually the least costly alternative.

We should note that a massive change in the role of assessment has occurred in the last two decades. The measurements (tests and so on) have had their usability nominally transformed. Instead of just showing (albeit approximately) how well students and organizations perform, with the understanding that the performance is merely an estimate or a sample of what people might do, given longer times, greater flexibility, or a slightly different problem set, now the scores on the test have become ends in themselves. There is a widespread belief that there is no better way to measure learning than to obtain a test score, any test score, and that good things should always be inferred from test scores going up, and bad things from test scores going down. As a result, in many sectors, a second-stage transformation has occurred that legitimates only those activities congruent with the test content and form, with a result that topics or areas that are untested also remain untaught. And so, policymakers have made what was the *measurement* into the educational *intervention*. Policymakers have embraced this notion, thinking that if we had good tests, then this natural focus would serve both children and the educational systems in which they participate.

Additional justification for the use of test performance as the key measure of quality is the evidence that economic success is somewhat connected to prior test performance. For regions, states, or countries in competition with one another, attracting families and businesses with high test scores has become part of the regular sales pitch for real estate people, with the result that children will be with children like themselves, at least on an academic measure.

So we have, as some might say, a nifty system. A relatively cheap set of instruments allows us to distinguish among people and among institutions, on the basis of proficiency of some sort, and the rewards and sanctions that follow (known as accountability) have made performance on these instruments paramount. What could possibly be wrong with this approach? One way to pose the question is this: Are assessment results generating unusable, usable, or useful knowledge? Or in other words, are assessments solving problems directly related to learning?

Let's take a moment to consider what the assessment of learning is all about, how it may work, and how it often does work. Then we will apply to assessment the standards of unusable, usable, or useful knowledge.

Assessment is a feature of all learning, even if it is only the informal self-questioning that occurs when one has patiently turned book leaves and realizes that not one idea has been remembered. Assessment is also the process in which tender inquiries are made by teachers asking a child to explain why a particular strategy was proposed. More commonly, assessments are formal examinations, generated internally or externally, and are usually intended to hold consequences for the examinee. These consequences may involve receiving a teacher's mark or grade, a score allowing access to a particular course, a diploma with or without endorsement, or on the downside, a signal to reconfigure plans and hopes. Assessment in its purest form gives feedback, and the more adapted the assessment is to what the learner is experiencing and the capacities and learnings brought into the assessment situation, the more likely the assessment will promote growth and accomplishment. It will be useful. At least that is the story.

Reality stands apart, unfortunately. Assessment data, when used on a large scale for system monitoring purposes, may be neither adaptive nor appropriate to the learner, and may not provide information that can be used by the student or the teacher. Some assessments may likely conflict with other data and create a problem in understanding what it all means. Furthermore, assessment data may not be in a form that is easily digested. Moreover, because formal assessments have a quantitative flavor—statistical transformations and the like—they exude science, and that alone may swamp the credibility of other sources of assessment knowledge. We must get inside the makings of tests to see how they relate to learning and what sense can really be made from data rather than adopting a passive position and assuming that experts have done the thing right.

What types of knowledge about assessment itself do we really need as educators? I claim that we first must know the purpose, or purposes, of one or more assessments (tests). Who is really the primary audience? Who is to make use of the information, and what decisions will hinge on it? These questions are rapidly followed by questions of what to assess, who gets examined and how frequently, how to design assessments, how to interpret results, and how to determine whether the results are to be trusted, in the light of the purposes they serve (or validity). Are examinations evaluating a particular reading program appropriate to gauge school

effectiveness? Should university admissions test scores (as in the United States) be used to judge the quality of secondary schools?

To address this set of Problems, I believe that we must make sure that the assessments given for *whatever purpose*, whether large scale or in the classroom, serve first the learning and the learners. At the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), we have been working for 15 years or so on a strategy to design and implement assessments so that they meet three criteria: (a) They lead to coherent, sustained learning; (b) they support a spiral form of teaching, each enhancing and linking what has come before; and (c) they direct students to knowledge and skills that can be transferred (in psychological terms) or applied to new or unforeseen situations. The formats in or authorities under which the assessments are given are less important than the learning they actually stimulate.

In brief, CRESST models are based on research knowledge. To design a test or assessment, we first focus on desired student cognition and learning. We then reverse the usual way tests are built: Instead of starting with subject matter—world history, for instance—we begin with the cognitive expectations—are we focusing on communication, content understanding, problem solving, or some combination? After deciding on the family of cognitive demands, we use a template (representing a translation of research into a usable form). Then, we return to the subject matter domain and apply a template or structure of the assessment model to it, substituting content or examples as needed. In history, we could address Asia in the 19th century, the Spanish expansion, or the complexities of the Cold War using the same general model and specific template. This approach forces a level of coherence among sets of assessment tasks, among subject matters, and among authorities (bureaucratic levels) administering the test. It allows teachers to "line up" their instruction and formative assessments with external mandates without corrupting the tests or their view of teaching. It also supports vertical integration from grade to grade. It saves money because it allows important task architecture to be reused. While this isn't the forum for detailed procedures, let me provide a brief sketch of how this works.

Here is a representation of the approach, with learning in the middle of five key families of intellectual skills or cognitive demands: content understanding, problem solving, communication, metacognition (or actively controlling your own

learning) and teamwork and collaboration (Figure 1). I'll illustrate the content understanding model in part.

First we figured out what the key elements are, from research, that describe significant content understanding. Accept for the moment that they are (a) understanding the big ideas in a domain, (b) seeing their relationships, (c) avoiding misconceptions, and (d) using prior knowledge and resources to convey meaning. In content understanding, I'll illustrate two examples of templates, drawn from the model, that produce different looking tasks but share the same deep infrastructure. We at CRESST (with teacher advice) have agreed that we want children to read or encounter real text, or representations of artifacts, whether historical or current, or literary, scientific, or artistic. Thus the specification for the task requires the presentation of primary source materials. We also need—as the research supports—students to demonstrate that they can integrate specific prior knowledge with higher principles or themes. This process of translating research into models, models into templates, and templates into coherent assessments represents our strategy for making research knowledge "usable." We have decided, again based on lots of cognitive research (Chi, Glaser, & Farr, 1988) and a modicum of logic, that students' work should be scored on models based on experts' performance rather than using an abstract idea of what "good" work is.

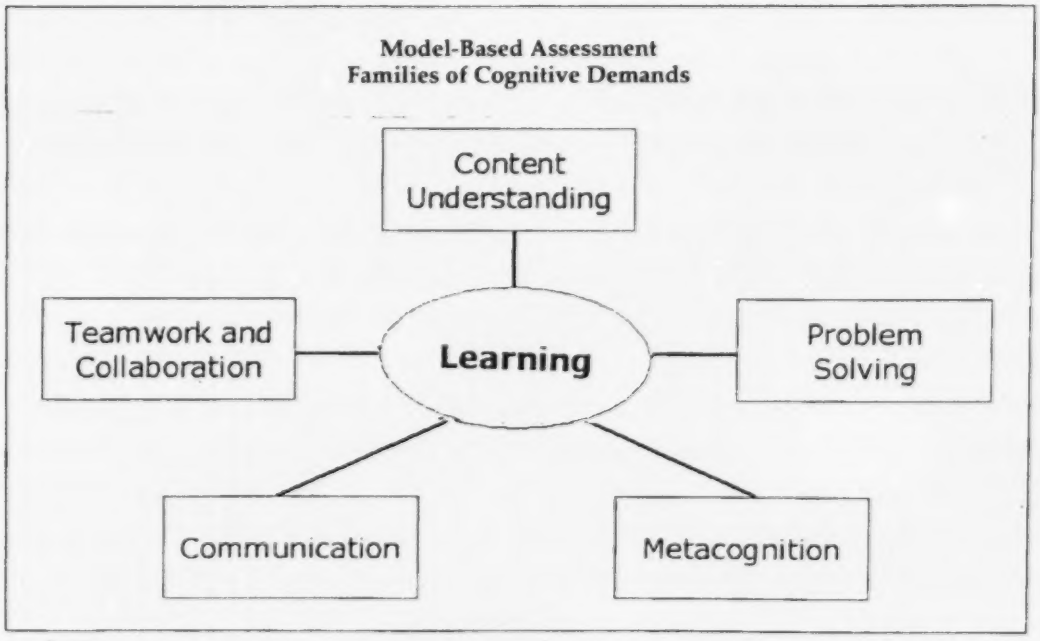


Figure 1. Model-based assessment example.

BEST COPY AVAILABLE

Each cognitive model can generate multiple templates. In the first example of a template for content understanding (Figure 2), students are given a writing task in history, and after reading primary source materials of considerable length, they construct an answer evaluated by a scoring rubric based on expert performance. In an example in Hawaiian history, students first read instructions for the task (Figure 3). Figure 4 shows an excerpt of one of the longer documents the students (12-year-olds) would read. The scoring rubric is shown in Figure 5. We have used this framework in Grades 2 through university (with appropriate modifications), and in subject matters ranging from chemistry to humanities to mathematics. Now remember, the model is about deep understanding. The second template to help generate multiple assessment tasks asks students to use (usually on a computer) a graphical task to show relationships within a domain (Figure 6). There are a number of ways, sometimes more than one in any student's work, to organize a field. In this second history example, students were given primary source materials to read (writings of a Depression era United States president and his opponent) and asked to map their understanding (Figure 7). The same approach has been used in secondary school genetics and in an adult literacy measure. These representations, just as in the template for content understanding (Figure 2), are scored by using experts' responses to the questions. The cognitive demands of the tasks are similar, even though the formats of the tasks differ. The relationship between the written and graphic tasks is about .6—approximately the same relationship as between parents' education level and students' achievement. Making research knowledge usable required our abstracting from the fields of learning, psychometrics, and psychology, conducting some of our own studies, and trying out the approach on a small scale, statewide, and as a regular part now of the annual assessments of 500,000 children.

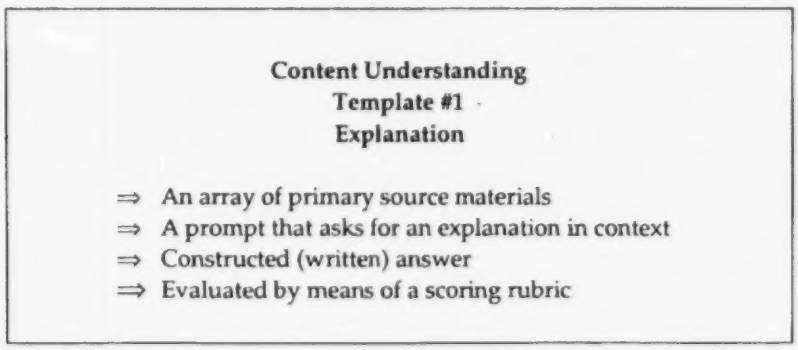

Content Understanding
Template #1
Explanation
⇒ An array of primary source materials
⇒ A prompt that asks for an explanation in context
⇒ Constructed (written) answer
⇒ Evaluated by means of a scoring rubric

Figure 2. Template 1—Explanation example.

Hawaiian History Writing Assignment:
Bayonet Constitution

Imagine you are in a class that has been studying Hawaiian history. One of your friends, who is a new student in the class, has missed all the classes. Recently, your class began studying the Bayonet Constitution. Your friend is very interested in this topic and asks you to explain everything that you have learned about it.

Write an essay explaining the most important ideas you want your friend to understand. Include what you have already learned in class about Hawaiian history, and what you have learned from the texts you have just read. While you write, think about what Thurston and Liliuokalani said about the Bayonet Constitution, and what is shown in the other materials.

Your essay should be based on two major sources:

1. The general concepts and specific facts you know about Hawaiian history, and especially what you know about the period of the Bayonet Constitution.
2. What you have learned from the readings yesterday.

Be sure to show the relationships among your ideas and facts.

Figure 3. Hawaiian history writing assignment example.

Excerpts from Hawaiian History
Primary Source Documents

Liliuokalani

For many years our sovereigns had welcomed the advice of American residents who had established industries on the islands. As they became wealthy, their greed and their love of power increased. Although settled among us, and drawing their wealth from resources, they were alien to us in their customs and ideas, and desired above all things to secure their own personal benefit.

Kalakaua valued the commercial and industrial prosperity of his kingdom highly. He sought honestly to secure it for every class of people, alien or native. Kalakaua's highest desire was to be a true sovereign, the chief servant of a happy, prosperous, and progressive people.

And now, without any provocation on the part of the king, having matured their plans in secret, the men of foreign birth rose one day *en masse*, called a public meeting, and forced the king to sign a constitution of their own preparation, a document which deprived [him] of all power and practically took away the franchise from the Hawaiian race.

Figure 4. Hawaiian history primary source document example.

History Explanation
Scoring Rubric

⇒ General impression of content quality
⇒ Principles or concepts
⇒ Prior knowledge
⇒ Use of available resources
⇒ Misconceptions (negative)
⇒ Argumentation (domain appropriate)
⇒ English mechanics

Figure 5. History explanation scoring rubric example.

Content Understanding
Template #2
Knowledge Representation

⇒ Key aspects of ideas, supporting facts and views and their relationships
⇒ Relationship is explicit
⇒ Organizational options
  - o Core and peripheral
  - o Hierarchical
  - o Cause and effect
  - o Chronological

⇒ Expert scoring

Figure 6. Template 2—Knowledge representation example.

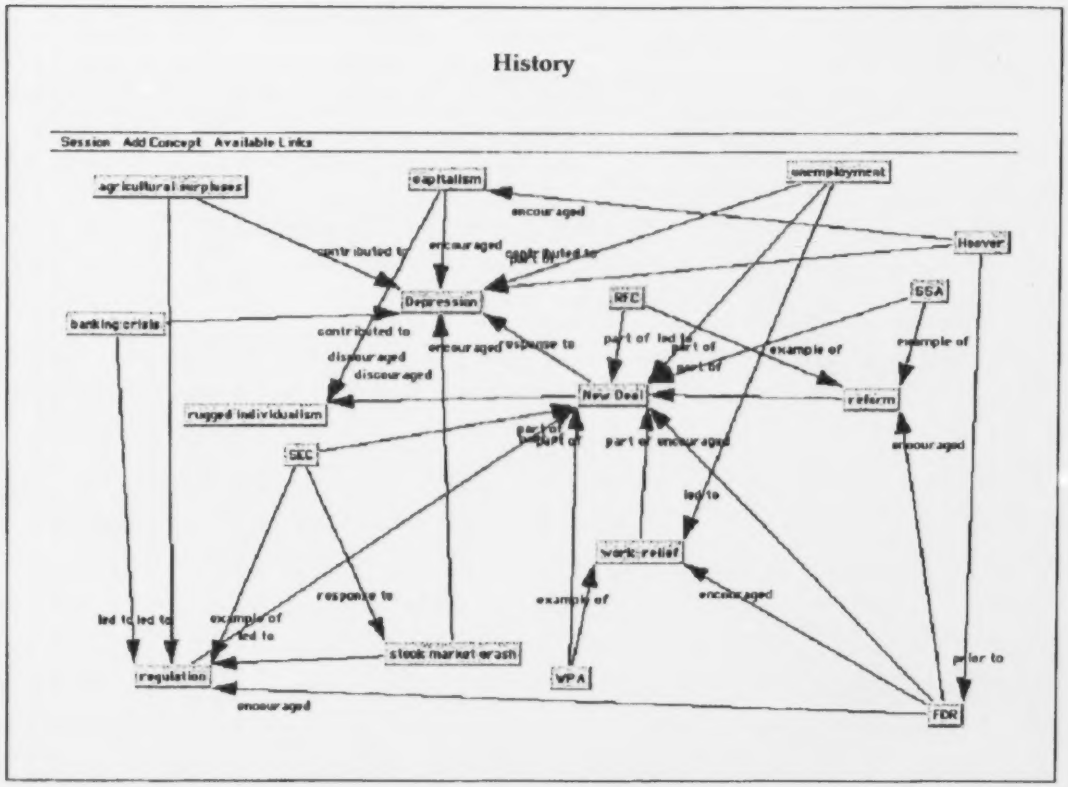


Figure 7. History mapper example.

To take the next step, and transform knowledge into a useful form, assessment itself is not sufficient. Assessments need to be timed so that the data can be used. Those teaching and those expected to make interpretations should possess high levels of content knowledge (in fact, teachers' self-report of their own topical knowledge has recurrently shown up as a big predictor of student performance on complex academic tasks, at elementary, middle, and secondary school levels). But we can help make assessment knowledge useful by providing help that ensures teachers and other educators know as much as they can about their students. In systems where good records are kept, teachers and other instructional leaders have to learn how to combine test results and other sources of information, and how to weigh or value different information. When performance falls short of expectations or requirements, teachers need to know where to find help, and they need assistance in knowing what to do (too often they may fall back on a failed method). These last two points extend beyond the assessment remit, but are related to the careful documentation of the models guiding assessment design. Teachers can see whether

BEST COPY AVAILABLE

children performed poorly because they didn't have sufficient prior knowledge, had difficulty integrating new and old information, or perhaps could not organize their thoughts into more global principles.

Interpreting Results

One area that both teachers and school managers have little experience with (at least in the United States) is serious interpretation of assessment results. Questions need to be answered about how good the results are (in comparison to what), how to integrate classroom and other sources of information, and how to think reflectively in order to infer a reasonable next step.

One approach is to use what business calls decision support systems. These are software that allows easy query and manipulation of data. They work in part like browsers and in part like spreadsheets. The problem with many of these systems is that they are not sensitive to educators' needs. CRESST has tried again to translate research knowledge into tools that help make this type of analysis easy, productive, and even fun. These systems can meet one of the biggest challenges, that of incoherent information, allowing the identification of conflicting or similar data among different sets of students, tests by kind of task, subject matter, or instructional history. These systems are clearly adjuncts for the teacher, rather than machines that spit out right answers. They depend upon the insightful question that a good teacher may think to ask in order to explain information. CRESST has created the Quality School Portfolio (QSP), originally just to show a prototype of what could be done. QSP is used in more than 1,000 schools and in every state in the U.S. It has been transformed into a Web-based system and is currently being tried in states with varying kinds of accountability systems, including Illinois, Indiana, Missouri, Nebraska, New Hampshire, and New Jersey. Following these pilots, emphasizing the "webbiness" of the system and the classroom and parent interface, the system will go nationwide, for free. The system components have local, school, classroom, and parent functions. There is also a place to access student work. Our studies suggest that teachers and principals find great value in the system, especially those with little external support for data analysis. The creation of an individual record for a student is a boon for teacher, student, and parent. Again, we have tried to transform data (some of it usable) into a more useful form.

Utility Is Context Dependent

We believe that researchers can go a long way to help make their findings more usable—at least capable of being understood and tried in a variety of settings. Providing tools such as assessment templates (or soon-to-be-released assessment authoring systems) will help teachers by raising the quality of some of what they do without raising the time expenditure commensurately.

But the hard part of knowledge-based reform is both general and specific. A fundamental change is required of many teachers—a shift from a chronological perspective of what I will do Monday, or in March, to what should each learner be doing. Such a cultural shift needs leadership, tools, time, and collaboration to succeed. Moreover, it requires that the administration of schools, at all levels, be willing to take a chance on change, and be ready to revise if well-thought-out plans fail to yield results. The context for success of knowledge-based reform is key. Knowledge must be locally owned and valued, and the infrastructure must allow enough stability for trials. Staffs need the capacity to investigate, including time and tools. Learning must be the major outcome, and where differences exist between local and external policies, a way to reach congruence or a temporal peace must be pursued from all sides.

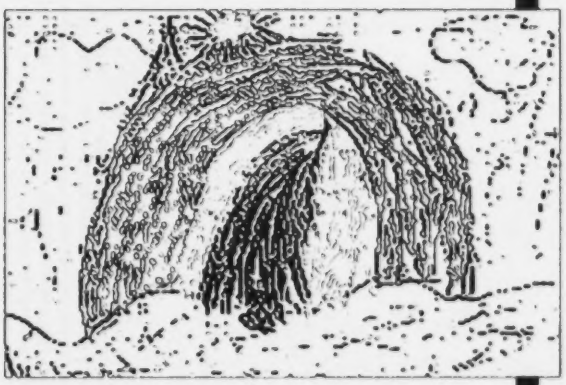
Whether knowledge is useful, of course, depends on the pudding. In other words, research and assessment and data interpretation form part of the foundation for change. For assessment knowledge and results to be useful, context, capacity, and communication of the teaching and learning system are key. Unless assessment knowledge is ultimately useful to students who do the learning, it is no more than a comforting management exercise. Useful knowledge must go to the heart of why, what, and how students learn.

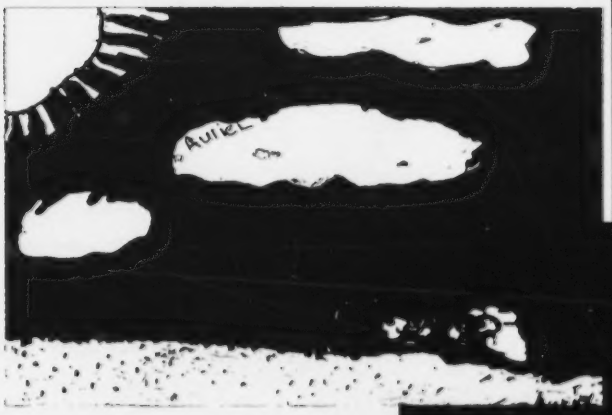
References

Chi, M. T. H., Glaser, R., & Farr, M. (Eds.). (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Lindblom, C. E., & Cohen, D. K. (1979). *Usable knowledge: Social science and social problem solving*. New Haven, CT: Yale University Press.

Weiss, C. H. (1977). Research for policy's sake: The enlightenment function of social research. *Policy Analysis, 3,* 531-45.





BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").